

ADDITIONAL	WHITE SECTION <input checked="" type="checkbox"/>
CESTI	BLUE SECTION <input type="checkbox"/>
SDC	GREEN SECTION <input type="checkbox"/>
UNCLASSIFIED	<input type="checkbox"/>
DATE	14 73
BY	LM
DISTRIBUTION AVAILABILITY CODES	
DIST.	AVAIL. CODE SPECIAL
1	

SP-2556/000/00

SP a professional paper

GUIDELINES FOR THE UTILIZATION OF
STATISTICIANS IN THE DESIGN
AND EXECUTION OF INFORMATION
RETRIEVAL SYSTEM EVALUATION STUDIES

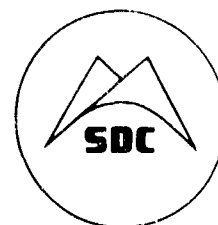
by

Alan A. Anderson
System Development Corporation
Dayton, Ohio, U.S.A.

12 July 1966

SYSTEM
DEVELOPMENT
CORPORATION
2500 COLORADO AVE.
SANTA MONICA
CALIFORNIA
90406

International Advanced Study Institute on
the Evaluation of Information Retrieval
Systems Proceedings at The Hague, Netherlands
20 July 1965.



12 July 1966

1
(Page 2 Blank)

SP-2556/000/00

ABSTRACT

Expressed desires of lecturers and members at the NATO Advanced Study Institute, July 1965, led the author to prepare guidelines for using statisticians in retrieval system evaluation studies. These guidelines describe questions by statisticians for information concerning the retrieval system environment. Specific statistician tasks are identified. The user support requirements are outlined. Constraints in the utilization of a statistician are discussed. The paper was conceived, reproduced, written, and distributed among the institute's two-week session.

GLOSSARY

1. Criterion Measure = A scale which represents an acceptable (to management) standard against which all other measures of system performance are evaluated by analytical techniques. For example, document value to the user is an underlying concept for a criterion measure.
2. Analysis Units = Data points. That is, a sample number with a combination of characteristics which defines a manipulatable set of information.
3. Objectivity = The quality or state of being verifiable by scientific methods.
4. Reliability = The quality or state of being consistent upon repeated measurement or observation.
5. Validity = The quality or state of being correlated or in harmony with the criterion.
6. Subjective Data = Observations of system behavior phenomena which involve human perception, including opinions, ratings, etc.
7. Parametric Statistics = A set of statistical procedures which specifies rigid requirements for measurement and sampling, e.g. interval scales, large samples, etc.
8. Degrees of Freedom = The number of observations which are free to vary after certain restrictions have been placed on the data by the design of the analysis or choice of statistical tool.
9. Cross-validation = The verification of the results of an initial evaluation by a second analysis of data with a different sample using the same variables.
10. Frame of Reference = A systematic set of principles, rules, or presuppositions or a system of laws, mores, or values, or an interlocking group of facts or ideas serving to orient or give particular meaning (as to fact, statement, or point of view), or serving as a matrix for behavior or the formation of attitudes

TABLE OF CONTENTS

<u>Paragraph</u>	<u>Page</u>
Glossary -----	3
Abstract -----	1
1.0 INTRODUCTION -----	5
2.0 QUESTIONS ASKED BY THE STATISTICIAN -----	5
2.1 What Is the Problem? -----	5
2.2 What Data Is Available or Can Be Obtained? -----	5
2.3 What Is the Criterion? -----	5
2.4 What Resources Are Available for Data Gathering? -----	5
2.5 What Resources Are Available for Data Analysis? -----	6
2.6 What Subjects Are Available (if required)? -----	6
2.7 How Much Calendar Time Is Available for Data Gathering and Analysis? -----	7
2.8 What Results of Prior Studies on the Problem or Related Problems Are Available? -----	7
3.0 WHAT THE STATISTICIAN WILL DO -----	7
3.1 Identify Basic Assumptions -----	7
3.2 Define the Variables and Analysis Units -----	7
3.3 Determine Analysis Strategy -----	7
3.4 Provide Logical Basis for Interpreting the Results Including Generalizing Limitations -----	9
3.5 Determine the Degree of Definitiveness Obtainable -----	9
3.6 Relate Results to Prior Study Results -----	9
3.7 Author or Co-Author Final Report(s) -----	9
3.8 Estimate Costs -----	9
4.0 FROM THE USER SUPPORT FOR THE STATISTICIAN IS NEEDED IN THE FOLLOWING: -----	10
5.0 CONSTRAINTS IN THE UTILIZATION OF A STATISTICIAN -----	10
5.1 Unfamiliarity with the System -----	10
5.2 Short Supply of Adequate Statisticians -----	11
5.3 Order of Desirability in Statistician Types -----	11
5.4 Contractor or Staff Member -----	11

1.0 Introduction

The need for statistical support in designing and executing evaluation studies became more clear as the NATO Advanced Study Institute conducted its proceedings. Several lecturers and members expressed a desire for help in using this kind of support.

Informal discussions revealed an interest in having guidelines for the utilization of statisticians as a part of the report on the proceedings of this Institute. In response to this need the author formulated this preliminary draft as proposed guidelines.

2.0 Questions Asked by the Statistician

The statistician will need to have answers to certain kinds of questions before he can begin to perform his job. These questions are based on the information which only the system-user knows. Typical questions are as follows:

2.1 What Is the Problem?

The main information that the statistician needs to know is a clear description of the problem. This should include the objectives, the conditions, the requirements and the constraints in a summary of the system, which should cover all these aspects of the problem. Later, questions are described in detail. A clear understanding of the problem between the user and the statistician will have an important bearing on the effectiveness of the statistician's performance.

2.2 What Data Is Available or Can Be Obtained?

The first specific information the statistician needs to know is an objective description of the kind of data to be used in the analysis. At this point it is not necessary to define the variables. That is the statistician's job at a later stage of the design development.

2.3 What Is the Criterion?

The most important variable is the criterion measure, so early in the discussions between the user and the statistician the logic of the criterion measure must be determined.

2.4 What Resources Are Available for Data Gathering?

The design of the study must take into account the resources which are available for data gathering. Will the statistician acquire

data? Will the user supply all the data? What communication media are required for gathering data? Are data already available? How much additional data are required? These are questions which must be answered.

2.5 What Resources Are Available for Data Analysis?

The analysis of data usually requires three categories of resources; they are: clerical, professional and machine or equipment.

2.5.1 Clerical

For clerical assistance three types are usually needed:

1) data recording 2) editing capabilities and 3) statistical capabilities.

2.5.2 Professional

If the data are complex and require a comprehensive knowledge for certain steps in processing, professional capabilities are required. Usually one or more subject matter specialists is necessary for this part of data preparation and analysis. In addition, if complex computer programming is required, a professional programmer is necessary. Even though computer programs are available, the services, on a part-time basis, of a professional programmer will enhance the quality of the study.

2.5.3 Machine or Equipment

The volume of data, the complexity of the design, and the number of steps and processes are important factors which determine machine or equipment requirements. The requirements may vary from the use of a desk-calculator to complex electrical accounting machines or possibly computers or specialized equipment.

The machine requirements can represent a considerable financial cost, but at the same time bring about a great saving in time and increase the meaningfulness of the output. Machine requirements are directly proportional to the number of variables, size of sample data, and the complexity of analytical techniques.

2.6 What Subjects Are Available (if required)?

If judgments or ratings are to be required by professional subjects or other types of subjects in developing criterion and/or reference

measures, an early assessment of these requirements is critical. Plans should be initiated early to arrange for such requirements in subjects.

2.7 How Much Calendar Time Is Available for Data Gathering and Analysis?

As prior questions are answered the amount of calendar time to accomplish the requirements will tend to be larger than the user anticipated. A review of all requirements with reference to the possible available calendar time will cause a revision in objectives and other study design conditions.

2.8 What Results of Prior Studies on the Problem or Related Problems Are Available?

Before the final design is established it is important that the statistician have all available results of prior studies or activities which are related to the problem.

3.0 What the Statistician Will Do

After the statistician has had an opportunity to study and evaluate discussions with the user on the questions described in the previous section he will be prepared to perform his assignment.

3.1 Identify Basic Assumptions

On the basis of prior studies and insight into the nature of the problem, basic assumptions are identified by the statistician.

3.2 Define the Variables and Analysis Units

On the basis of the basic assumptions the variables and analysis units are to be defined by the statistician.

3.3 Determine Analysis Strategy

At this point the crucial parts of the study design are formulated by the statistician. These include objectivity and reliability controls, statistical tool requirements, degrees of freedom requirements, estimates of expected validity and reliability levels, and any special features of the design which are implied in the problem and objectives.

A. Determine Objectivity and Reliability Controls

At this point the sampling design comes under consideration in addition to the methods for the gathering of subjective data.

If a larger part of the data are subjective in nature, a large number of control variables are necessary for properly assessing the objectivity of the data. The validity of the results requires testing, and such tests should be provided for in the sampling design. For example fortuitous effects in the data may show favorable results in a single sample. Such effects may not exist in another random sample. Consequently, conclusions based on a single sample, are risky. It may be necessary to add additional control variables at this point.

B. Determine Statistical Tool Requirements (Parametric or Non-Parametric).

Literature on statistics contain a large number of useful tools for analysis. The law of parsimony should be applied to the selection of statistical tools. Sophisticated tools should never be used merely because they are available or because a statistician knows how to use them. In some cases specialized techniques or new applications of existing tools may be required. In rare cases novel techniques or tools will be required. The development of novel techniques is usually very costly.

C. Determine the Degrees of Freedom Requirements

The number of degrees of freedom which are required are determined by a combination of factors, such as the size of the population or universe, the complexity of the data in terms of numbers of variables or data characteristics, the degree of subjectivity in the data and the complexity of the analysis design. The number of degrees of freedom is also known as the sample size or the number of cases required for the analysis. At this point the objectivity, reliability and validity controls should be taken into account. Provision should be made for adequate cross-validation of results.

D. Estimate the Validity Level

At this point a statistician should have sufficient knowledge of the problem, design, objectives, and the data to make an estimate of the level of validity which can be obtained. In cases where a problem requires a pilot study design this requirement will be too difficult.

F. Estimate the Reliability Level

In the case of reliability the experienced statistician should be able to make a reasonable estimate of the expected reliability

for the criterion measures and other variables.

3.4 Provide Logical Basis for Interpreting the Results Including Generalizing Limitations.

On the basis of the review of prior results and/or agreement with the user of the system the statistician formulates a frame of reference for interpreting the data. He should also evaluate the context of the problem and assess the extent to which the expected results can be applied or generalized to related problems.

3.5 Determine the Degree of Definitiveness Obtainable

The statistician will probably be overcautious in claiming that definitive results will be obtained. However, the user will want to know how much dependence he can place in the expected results in order to justify the costs. Therefore the statistician is required to "stick his neck out".

3.6 Relate Results to Prior Study Results

As a part of the design for this particular study the statistician should select tabular and other statistical material which may be incorporated in the report for the study in question. These data from prior reports should also be considered under item D above.

3.7 Author or Co-Author Final Report(s)

Agreement should be reached between the user of the system, including representative professional personnel, and the statistician concerning responsibility for preparing final reports and author credits.

3.8 Estimate Costs

A final but not least important consideration is the matter of financial and other costs. These can be usually defined as calendar time requirements, personnel requirements and machine or equipment requirements.

A. Temporal Requirements

After all the previous steps have been accomplished a realistic date for completion of the project should be established through negotiation between the user and the statistician. This date should allow for "Slippage".

B. Personnel Requirements

A final assessment of personnel for data gathering, analysis, travel, and report writing is to be made.

C. Equipment Requirements

Last but not least, in terms of financial costs, is the cost of equipment requirements for the analysis of data.

D. Overhead Requirements

In the process of executing of the study unexpected costs always arise, consequently provision should be made for enough overhead or fixed fees which will insure support for the final completion of the project.

4.0 From the User Support for the Statistician is Needed in the Following:

- A. Identifying Basic Assumptions.
- B. Defining Variables.
- C. Arranging for Subjects (if required).
- D. Arranging for Data Gathering.
- E. Arranging for Data Analysis:

- 1. Clerical
- 2. Professional
- 3. Machine or Equipment

- F. Implementing Data Analysis.

5.0 Constraints in the Utilization of a Statistician

Although the need for a statistician is very important the user of a system should recognize the constraints which effect the use of this kind of professional help in evaluation studies. Among these constraints are the general unfamiliarity of a new statistician with the user's system, the short supply of adequate statisticians and especially the difficulty of obtaining the most desirable type.

5.1 Unfamiliarity with the System

System environment, personnel, conventions and type of information involved represent a complex which can only be adequately comprehended through long experience.

5.2 Short Supply of Adequate Statisticians

Statisticians are in great demand throughout the entire realm of all professional activities. The nature of the work is difficult and adequate competence is not easy to obtain. Therefore other types of professionals are usually easier to find than statisticians. In addition the cost of statisticians is high, because of their market value as a function of the law of supply and demand.

5.3 Order of Desirability in Statistician Types

1. Psychometrician at the Ph.D. or M.A. level with three or more years in research design and analysis experience and three or more years of data processing experience including clerical, EAM, and digital computer.
2. Econometrician, sociometrician or biometrician with the same qualifications as above.
3. Statistician in any other subject matter field with the same qualifications as above.

5.4 Contractor or Staff Member

The user should carefully consider the three possible alternatives in providing for statistical design support; i.e. contractor, full time employee, or part-time employee. If the project is very large and relatively short in duration, a contractor may be appropriate. If the problem is reasonably large and continuous, a full time employee may be justified.

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) System Development Corporation Santa Monica, California		2a. REPORT SECURITY CLASSIFICATION Unclassified	
3. REPORT TITLE Guidelines for the Utilization of Statisticians in the Design and Execution of Information Retrieval System Evaluation Studies			
4. DESCRIPTIVE NOTES (Type of report and inclusive dates)			
5. AUTHOR(S) (Last name, first name, initial) Anderson, A. A.			
6. REPORT DATE 12 July 1966		7a. TOTAL NO. OF PAGES 10	7b. NO. OF REFS
8a. CONTRACT OR GRANT NO. Personal/ Professional		8a. ORIGINATOR'S REPORT NUMBER(S) SP-2556	
a. PROJECT NO.		8b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	
c.			
d.			
10. AVAILABILITY/LIMITATION NOTICES Distribution of this document is unlimited			
11. SUPPLEMENTARY NOTES		12. SPONSORING MILITARY ACTIVITY	
13. ABSTRACT Expressed desires of lecturers and members at the NATO Advanced Study Institute, July 1965, led the author to prepare guidelines for using statisticians in retrieval system evaluation studies. These guidelines describe questions by statisticians for information concerning the retrieval system environment. Specific statistician tasks are identified. The user support requirements are outlined. Constraints in the utilization of a statistician are discussed.			

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Information Storage & Retrieval Statistical Evaluation						

INSTRUCTIONS

1. **ORIGINATING ACTIVITY:** Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization (*corporate author*) issuing the report.

2a. **REPORT SECURITY CLASSIFICATION:** Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. **GROUP:** Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. **REPORT TITLE:** Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parentheses immediately following the title.

4. **DESCRIPTIVE NOTES:** If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. **AUTHOR(S):** Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. **REPORT DATE:** Enter the date of the report as day, month, year, or month, year. If more than one date appears on the report, use date of publication.

7a. **TOTAL NUMBER OF PAGES:** The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. **NUMBER OF REFERENCES:** Enter the total number of references cited in the report.

8a. **CONTRACT OR GRANT NUMBER:** If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. **PROJECT NUMBER:** Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. **ORIGINATOR'S REPORT NUMBER(S):** Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. **OTHER REPORT NUMBER(S):** If the report has been assigned any other report numbers (either by the originator or by the sponsor), also enter this number(s).

10. **AVAILABILITY LIMITATION NOTICES:** Enter any limitations on further dissemination of the report, other than those

imposed by security classification, using standard statements such as:

- (1) "Qualified requesters may obtain copies of this report from DDC."
- (2) "Foreign announcement and dissemination of this report by DDC is not authorized."
- (3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through _____."
- (4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through _____."
- (5) "All distribution of this report is controlled. Qualified DDC users shall request through _____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. **SUPPLEMENTARY NOTES:** Use for additional explanatory notes.

12. **SPONSORING MILITARY ACTIVITY:** Enter the name of the departmental project office or laboratory sponsoring (paying for) the research and development. Include address.

13. **ABSTRACT:** Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as (TS), (S), (C), or (U).

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. **KEY WORDS:** Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.